1 Better Distortion with Distribution

There is a well known lowerbound for the distortion of embedding a metric space (X,d) into ℓ_1.

Theorem 1.1 *For any metric space (X,d) on n points, one has*

$$(X,d) \xhookrightarrow{\Omega(\log n)} \ell_1.$$

For ℓ_2 the lowerbound is still $\Omega(\log n)$ [1].

Recall that we want to find a $(O(k),(1+\varepsilon)c)$-outlier embedding into ℓ_2 for any metric space (X,d) which admits a (k,c)-outlier embedding into ℓ_2. If we can do this deterministically, we actually find an embedding of the outlier points into ℓ_2 ~~with distortion $O(k)$, which contradicts the lowerbound~~. This is not true! The $\log k$ factor is required by SDP and only expansion bound is needed. We do not have to bound the contraction part. However, maybe we can do $O(k)$ via embedding into some distribution of ℓ_2 metrics.

Expected distortion Let (X,d) be the original metric space and let $\mathcal{Y} = \{(Y_1,d_1),\ldots(Y_h,d_h)\}$ be a set of target spaces. Let π be a distribution of embeddings into $\mathcal{Y}$. To be more precise, for each target space (Y_i,d_i) we define an embedding $\alpha_i : X \to Y_i$ and define the probability of choosing this embedding to be p_i. The original metric space (X,d) embeds into π with distortion D if there is an $r > 0$ such that for all $x,y \in X$,

$$r \leq \frac{\mathrm{E}_{i\leftarrow\pi}[d_i(\alpha_i(x),\alpha_i(y))]}{d(x,y)} \leq Dr.$$

Note that if we compute the minimum D for all x,y pair and take the average, the resulting value is called the average distortion.[2] There is an embedding into ℓ_p with constant average distortion for arbitrary metric spaces, while maintaining the same worst case bound provided by Bourgain's theorem.

The outlier paper (SODA23) also embeds (X,d) into distribution. We call this kind of embeddings stochastic embedding.

Lemma 1.2 *Let π be a stochastic embedding into ℓ_p with expected expansion bound $\mathrm{E}_{i\leftarrow\pi}\|\alpha_i(x)-\alpha_i(y)\|_p \leq c_{\mathrm{E}}d(x,y)$. Then there is a deterministic embedding into ℓ_p with the same expansion bound.*

Proof: We define a new averaged embedding $\alpha^*(x) = \sum_{i\leftarrow\pi}\alpha_i(x)p_i$. Consider the expansion bound for α^*.

$$\begin{aligned}
\|\alpha^*(x)-\alpha^*(y)\|_p &= \left\|\sum_{i\leftarrow\pi} p_i(\alpha_i(x)-\alpha_i(y))\right\|_p \\
&\leq \sum_{i\leftarrow\pi}\|p_i(\alpha_i(x)-\alpha_i(y))\|_p \\
&= \sum_{i\leftarrow\pi} p_i\|(\alpha_i(x)-\alpha_i(y))\|_p \\
&\leq c_{\mathrm{E}}d(x,y)
\end{aligned}$$

□

[1] https://web.stanford.edu/class/cs369m/cs369mlecture1.pdf
[2] https://www.cs.huji.ac.il/w~ittaia/papers/ABN-STOC06.pdf

Note that one cannot derive contraction bound for α^* from the stochastic embedding. So the distortion may not be the same.

Example: Random Trees Consider the problem of embedding some finite metric into a tree metric. We can get an $O(n)$ lowerbound via the unit edge length cycle C_n. However, if embedding into distortions is allowed, we can do $O(\log n)$.

Theorem 1.3 (Bartal) *Let (X,d) be a metric space on n points, let $\mathcal{D}T$ be the set of tree metrics that dominate d, there is a distribution π on $\mathcal{D}T$ such that (X,d) embeds into π with distortion $O(\log n)$.*

Is there any other known result on expected distortion of embeddings besides Bartal's theorem?

2 Stochastic Embedding into ℓ_2

We first ignore the outlier condition and see if stochastic embeddings break the $\Omega(\log n)$ lowerbound.

Theorem 2.1 (Bourgain) *For any metric space (X,d) and for any p, there is an embedding of (X,d) into $\ell_p^{O(\log^2 n)}$ with distortion $O(\log n)$.*

Bourgain develops a randomized algorithm that finds a desired embedding.[3] Can we get better expected distortion by repeating the algorithm and uniformly selecting an embedding? For the ℓ_2 case, the embedding has the following bounds:

1. Expansion. $\|f(x)-f(y)\|_2 \le O(\log n)d(x,y)$
2. Contraction. $\|f(x)-f(y)\|_2 \ge \frac{d(x,y)}{O(1)}$

The contraction bound is almost tight. Let K be the dimension of the target space. For the expansion bound, we have

$$\begin{aligned}\|f(x)-f(y)\|_2 &= \left(\sum_{i=1}^{K} |f_i(x)-f_i(y)|^2\right)^{1/2} \\ &\le \left(\sum_{i=1}^{K} d(x,y)^2\right)^{1/2} \\ &= \sqrt{K}d(x,y) \\ &= O(\log n)d(x,y)\end{aligned}$$

One thing we can try is to tighten the second line.

[3]The expansion bound always holds. The contraction bound holds with probability at least 1/2. See https://home.ttic.edu/~harry/teaching/pdf/lecture3.pdf

Bourgain's construction:
$m = 576 \log n$
for $j = 1$ to $\log n$:
  for $i = 1$ to m:
    choose set S_{ij} by sampling each node in X independently with probability 2^{-j}
    $f_{ij}(x) = \min_{s \in S_{ij}} d(x,s)$
$f(x) = \bigoplus_{j=1}^{\log n} \bigoplus_{i=1}^{m} f_{ij}(x)$ for all $x \in X$.

We want to show that for any fixed $x, y \in X$ and j,

$$\Pr[|f_{ij}(x) - f_{ij}(y)| \leq \frac{d(x,y)}{\operatorname{polylog} n}] \geq ???$$

One can see that our desired event does not happen with high probability for any pair of x, y. Let the original metric space be a line metric with n points. x, y locate on two endpoints of an interval and the rest $n-2$ points locate on the middle of xy. Then our metric in the target space $|f_{ij}(x) - f_{ij}(y)|$ is a $\operatorname{polylog} n$ factor smaller than $d(x,y)$ if and only if both x and y are selected in S_{ij}, which happens with probability 4^{-j}. This example shows that Bourgain's construction is tight up to a constant factor for some metric space.

3 Grid

Recall that we need an algorithm that outputs an embedding which extends a (k,c)-outlier embedding into ℓ_2 and we want the extended embedding to have a good (expected) expansion bound.

Conjecture 3.1 *Let (X,d) be a metric space such that $|X| = n$ and $\alpha : X \setminus K \to \mathbb{R}^d$ be a (k,c)-outlier embedding of (X,d) into ℓ_2^d, where $K \subseteq X$ is the outlier set. Then there exist an embedding $\beta : X \to \mathbb{R}^d$ such that β completes α and has expansion bound*

$$\max_{x,y \in X} \frac{\|\beta(x) - \beta(y)\|_2}{d(x,y)} \leq O(c\sqrt{\log k}).$$

In their bi-criteria approximation the dimension d is not important and therefore is considered as a fixed parameter. Conjecture 3.1 provides more tools than theorem 2.6, i.e. we know the coordinates of non-outlier points in the embedding β and we can use coordinates in $\mathbb{R}^d$ instead of simply mapping non-outlier points to outliers.

A common and powerful method is to use grid. We divide $\mathbb{R}^d$ into identical hypercubes of some sidelength s and working with grid cells instead of points. However, this method often involves the dimension d, which is not desirable...